Proskauer Rose LLP | 1001 Pennsylvania Avenue, NW | Washington, DC 20004

April 15, 2025

Re: *X Corp. v. Bright Data Ltd., 23-cv-03698 (N.D. Cal.)*[1]

Dear Judge Alsup:

Mark Twain famously said there are three types of lies: "lies, damn lies, and statistics." In opposing Bright Data's request for X's server data, X cross-moves for all Bright Data's activity logs beyond those it produced months ago. But allowing X to run its own queries on its systems using Bright Data's logs while denying Bright Data the ability to access the same data to cross-examine X is fundamentally unfair. Discovery is supposed to create an even playing field, not perpetuate an informational disadvantage. Bright Data opposes X's cross-motion because the Data Protocol that Bright Data proposes provides an orderly process for the mutual identification and production of relevant sever-related information.

Bright Data's Sample Logs Are an Appropriate Starting Point. Bright Data produced its full activity logs for two days as a *starting point* for discussions concerning the scope of data production. These logs contain the ***complete*** information regarding Bright Data's requests to the X platform on those days, including timestamps and the specific IP address (or proxy) used.[2] X does not complain about the content or completeness of the samples. Bright Data chose two non-controversial days—the day the Complaint was filed (July 26, 2023) and year-end (Dec. 31, 2023). And the proposed data protocol invites X to choose several days of its own. Starting with samples was not designed to limit log production to these days, but instead to show X what information Bright Data possesses, and to address X's position that it could not identify what relevant information X had until it sees Bright Data's logs. By producing these samples, Bright Data provided a path forward to the mutual identification and production of relevant server-related data.

X has had Bright Data's sample logs for over ***three*** months. Yet it has not identified any server failure or degraded user experience from Bright Data's conduct.[3] Last Thursday, X served Interrogatory responses conceding that it "cannot reasonably identify" any server failure and cannot "provide a response" based on these logs or any information it possesses. *See* Ex. 1.

X argues that it does not need to produce its information because the mere fact of server access – not its *effect* – is relevant. But this is not a *res ipsa loquitor* case. Specifically, X argues that "the key" is "exactly how, when, and to what extent" Bright Data communicated with X's servers, and that this is "information Bright Data – not X – should produce." ECF 247 at 1-2. Bright Data ***did*** produce this information. The fact that X still cannot show a server failure, or even any meaningful levels of access, shows those two days are not at issue in this case, and proves

[1] Emphasis added, internal citations and quotation marks omitted, and capitalizations conformed without brackets.

[2] X claims without support that "other products" are omitted. They are not. The logs Bright Data produced cover all traffic from Bright Data's network to an X or Twitter domain.

[3] It is no response that X cannot run Bright Data's IP addresses on its systems under the Protective Order. That is what experts are for. *See* Mar. 27, 2025 Hrg. Tr. 64 ("Now, you don't get to tell your in-house people what their secrets are. You have to do that through your experts."). Here, X can simply provide its experts its logs and server records for the two sample days, and X can produce the same data to Bright Data so its experts can do the same.

that not every log in Bright Data's possession is relevant. And they are only half of the equation. Bright Data is not asking X to take shots in the dark. X alleged instances of scraping in its Complaint, and it identified 9 instances (none by Bright Data) in its Interrogatory response. *See* Ex. 1. We need to start somewhere. Let's start there.

X Should Produce Its Side of the Communications. Like all conversations, server communications have two sides: sender and recipient. To prove an effect on the recipient, the recipient's information is critical. Yet X has refused to produce it. X says requiring Bright Data to produce logs for an additional 2,000 days will enable it to prove server effects when it has identified none from the sample already provided. Producing more will not cure X's deficiency.[4] The problem isn't the *amount* of Bright Data's logs; it is that Bright Data does not possess information about impact on X's servers. That information is in X's files.

X argues that producing its data is a "distract[ion.]" *Id*. In its view, its information is "ancillary-at-best" because "Bright Data and its customers are the *only* ones that know exactly" what parts of X's platform were accessed. *Id*. X knows that is not true. As Bright Data explained in its Interrogatory responses, Bright Data cannot identify the end points its proxy customers accessed because the communication between X and the customer is encrypted, meaning only X and the customer – not Bright Data – knows. Bright Data's sample logs confirm this. To draw an analogy, AT&T may know when you called your friend, but it does not know what you said. The same is true of Bright Data. Moreover, even if Bright Data knew what end-points were accessed, it would not know what servers were accessed, whether X owned those servers (as required for a trespass claim), or whether those servers were impaired.

In arguing otherwise, X draws an inapt analogy to other scraping cases, such as *hiQ Labs, Inc. v. LinkedIn Corp.*, 639 F. Supp. 3d 944 (N.D. Cal. 2022). But hiQ did not operate a patented proxy network, and the case was limited to hiQ's own scraping. When Bright Data itself scrapes, it does have information about the end-points it accessed. And it has produced this information, both in sample logs and the *actual* data itself. But unlike in *hiQ*, X also challenges ***third-parties'*** use of Bright Data's proxy network. Bright Data does ***not*** have information about the end-points its customers accessed. Only X has that information. Thus, information about which X end-points were accessed, which CAPTCHAs (if any) were encountered, what rate limits were triggered, how many requests were made (and their percentage of all traffic), and what effect they had on X's servers is all information in X's exclusive possession, custody, and control.

In resisting production of its logs, X argues it is enough to show the effects of automated access *generally*, but not from Bright Data *specifically*. To that end, X says it "will prove server harm through other documents," noting an "internal investigation memo,"[5] a so-called "Botox" analysis, some incident reports, a general description of its server architecture, and some emails from five of its twelve custodians (almost 90% of which were dumped last Thursday night). But

[4] X says it also needs the logs to "identi[fy] each customer scraping X." ECF 248 at 2. But Bright Data has now identified the potentially responsive customers through production of its CRM database.

[5] Though X did not produce metadata for these documents, their content suggests they were created in connection with X's *first* proposed Second Amended Complaint prepared by Quinn Emanuel (ECF 90-2 ¶¶ 70, 73). These allegations remain in the operative Second Amended Complaint (ECF 117-4 ¶¶ 87, 92). By relying on these documents, X failed to comply with its metadata production obligations, and waived privilege over its "investigation."

that does not preclude Bright Data's right to *disprove* X's claims through X's data. None of X's documents disclose any instance of server impairment due to Bright Data. Rather, these documents largely relate to types of bot activity that Bright Data does not engage in and is not even accused of. Nor do any of the documents contain systematic extracts of relevant information from X's data systems, which would allow Bright Data to show that its traffic was *de minimis*.

X's Burden Arguments Are Meritless. To avoid producing its data, X makes two contradictory arguments: the information does not exist, and it is too burdensome to produce. Neither has merit. Certainly, Bright Data is not asking for data that does not exist. Indeed, the purpose of the protocol is to explore what information does exist, and what types of analyses can be performed. Bright Data started that process by showing precisely the information contained in its logs. X should be required to produce its counter-part samples so the full universe of available information can be understood. Nor is it too burdensome to do so. It does not require X to produce *petabytes*, as X laments. Bright Data has not asked X to produce all logs relating to all requests by every person; it only asked for logs relating to specific IP addresses Bright Data used. Bright Data's logs were about 10 gigabytes, roughly a *million* times ***less*** than X's estimate. If Bright Data could produce these samples, there is no reason why X cannot do the same.

X Should Not Be Permitted to Manufacture One-Sided Evidence Using Bright Data's Logs. Nor is there merit to X's argument that it is "prohibitively" expensive to produce its data because there is "no user interface" and it would require engineering time. X's argument highlights the prejudice. In its view, X can run whatever analyses it wants, generate litigation-created summaries, and produce them as if they were ordinary course documents (which it has done). There is a name for that: manufacturing evidence. It is not permitted.

Bright Data needs the same access to the raw data as X. If X is unwilling to produce this information, the only fair way to proceed is through mutual samples. The parties can then each explore what information and analyses the other party's system is capable of generating, and then they can discuss how to ensure fair and equal access to such information. That is exactly what Bright Data's protocol accomplishes.

Requiring Bright Data to Produce Logs with Every IP Address Is Irrational, Disproportionate, and Prejudicial. Even aside from the fundamental unfairness of X's one-sided proposal, requiring Bright Data to produce over 2,000 days of logs raises both confidentiality and cost issues. Every IP address Bright Data discloses presents an opportunity for virtually *undetectable* abuse by X. X can introduce latency, or outright block, those addresses, without Bright Data being able to know the cause. Moreover, as Bright Data previously explained, Bright Data's active systems only contain 60 days of activity before it is archived. Restoring and hosting that information is costly. There is no reason for exposing Bright Data to this harm, or forcing Bright Data to incur this cost, if X cannot do anything with the data because its own retention policies prohibit it from doing any "IP-by-IP data" analysis. *See* ECF 247 at 2.

For these reasons, X's cross-motion to compel should be denied.

Sincerely,
/s/ Colin R. Kass
Counsel for Bright Data, Ltd.